Detecting Deceptive Hotel Reviews

using:
- Natural Language Processing
- Topic Modeling
- Machine Classification

SpringBoard – Capstone 2

11/2/2017 (Updated 2/9/2018)

Chris Gian

github.com/chrisgian

Roadmap

- Motivation
- Data / Data Prep / Variable Engineering
- Classification and Tuning
- Results
- Future steps

Deceptive Content Erodes Trust

KOTAKU

Assassin's Creed Origins Metacritic Flooded With Fake Positive User Reviews

HUFFPOST

ENTERTAINMENT 08/17/2016 06:10 pm ET

The Worst Of The Internet Is Sabotaging Amy Schumer's Book With Fake 1-Star Reviews

Can we not?

BUSINESS INSIDER

Tim Cook just outlined Apple's position on 'fake news'

Kif Leswing

17h 3,790

tripadvisor Playa del Carmen

Fake Hotel Reviews

Watch this Topic

Browse forums All | Riviera Maya forums | Playa del Carmen forum

Playa del Carmen Search

Toronto, On

Level 6 Contributor

297 posts

211 reviews

Fake Hotel Reviews

Save Topic

Mar 17, 2005, 12:19 PM

I have just read another, fake, vicious and malicious review for Hotel Playa Maya. It is dated March 15th and supposedly was posted by a Trip Advisor Member from England. It is SO OBVIOIUSLY FAKE - I can't believe that Trip Advisor even put it on the board. This troll is all over the place and Trip Advisor really must take some responsibility.

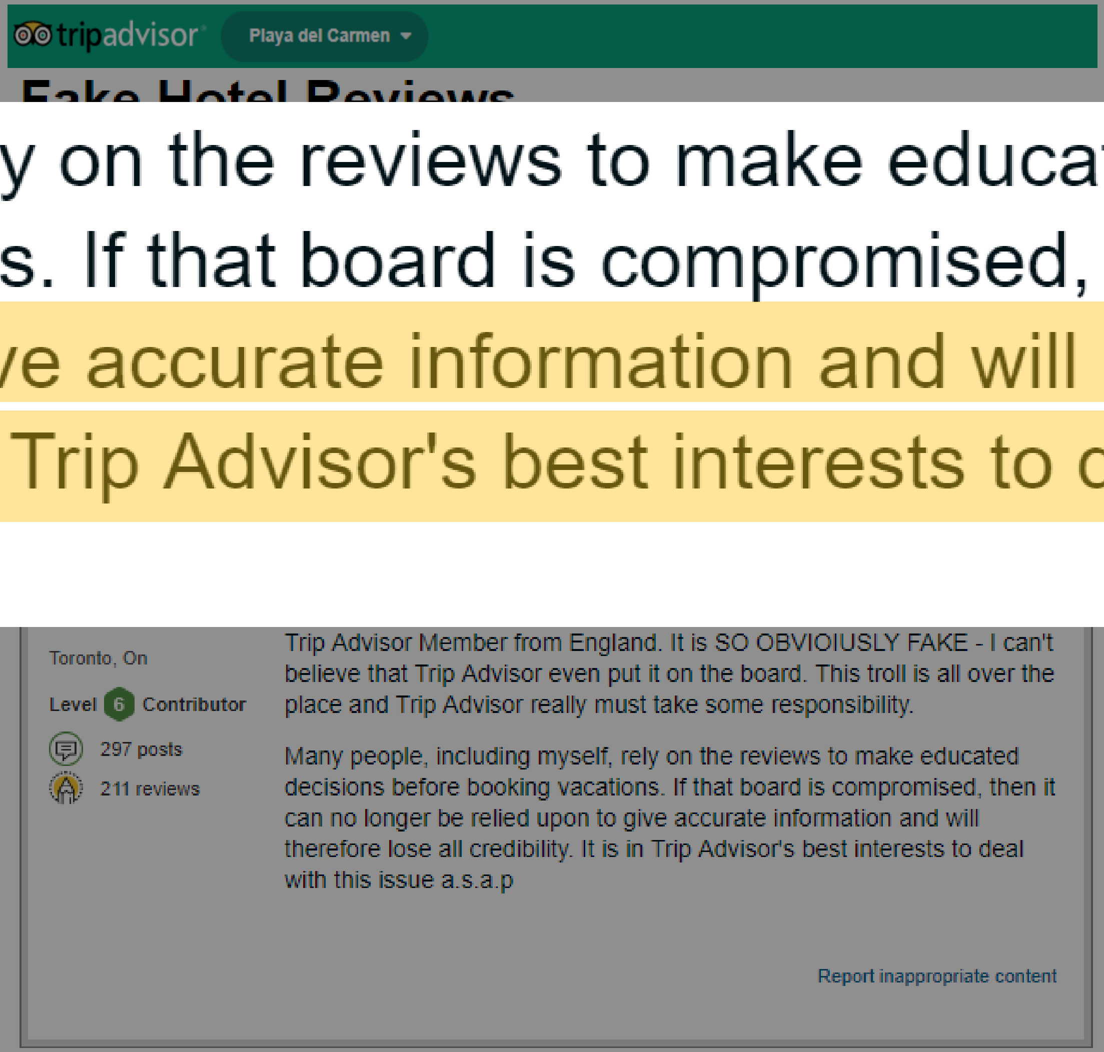
Many people, including myself, rely on the reviews to make educated decisions before booking vacations. If that board is compromised, then it can no longer be relied upon to give accurate information and will therefore lose all credibility. It is in Trip Advisor's best interests to deal with this issue a.s.a.p

Report inappropriate content

Deceptive Content Erodes Trust

KOTAKU

Many people, including myself, rely on the reviews to make educated decisions before booking vacations. If that board is compromised, then it can no longer be relied upon to give accurate information and will therefore lose all credibility. It is in Trip Advisor's best interests to deal with this issue a.s.a.p



tripadvisor Playa del Carmen

Fake Hotel Reviews

Toronto, On

Level 6 Contributor

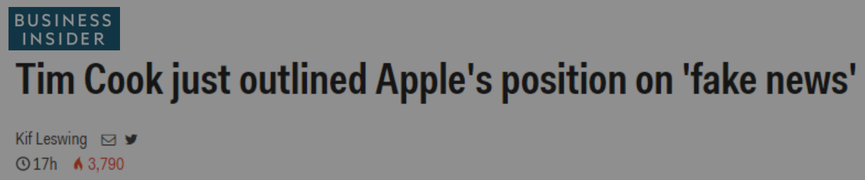
297 posts

211 reviews

Trip Advisor Member from England. It is SO OBVIOIUSLY FAKE - I can't believe that Trip Advisor even put it on the board. This troll is all over the place and Trip Advisor really must take some responsibility.

Many people, including myself, rely on the reviews to make educated decisions before booking vacations. If that board is compromised, then it can no longer be relied upon to give accurate information and will therefore lose all credibility. It is in Trip Advisor's best interests to deal with this issue a.s.a.p

Report inappropriate content



BUSINESS INSIDER

Tim Cook just outlined Apple's position on 'fake news'

Kif Leswing

17h 3,790

Can you tell the difference?

	Paid Reviewer	Real Reviewer
+	the experince at the hard rock hotel in chicago was fantastic,i will rate them a 6 out of 5. they have wonderful service and great staff and the view is just wonderful.	I stay at this hotel 2 times a year on business and LOVE it! The staff are great, the rooms are spacious and clean, and the location is perfect for shopping and dining on Michigan Ave.
-	The Swissotel Chicago is a very mediocre hotel, the service is always poor, and the room service food always comes cold, unless it's supposed to be cold than it comes warm. I would rather stay at a super 8 than this place again.	Uhhhhh, how do I know which direction to go in??? Stains on the carpet in the room. A big gauge in the wall, where maybe a thermostat once was??? The shower is decent. All in all, for the price they charge, NO THANKS!!! I'm just glad my company is footing the bill.

Examples shortened for space.

Hypothetical Client

Client:

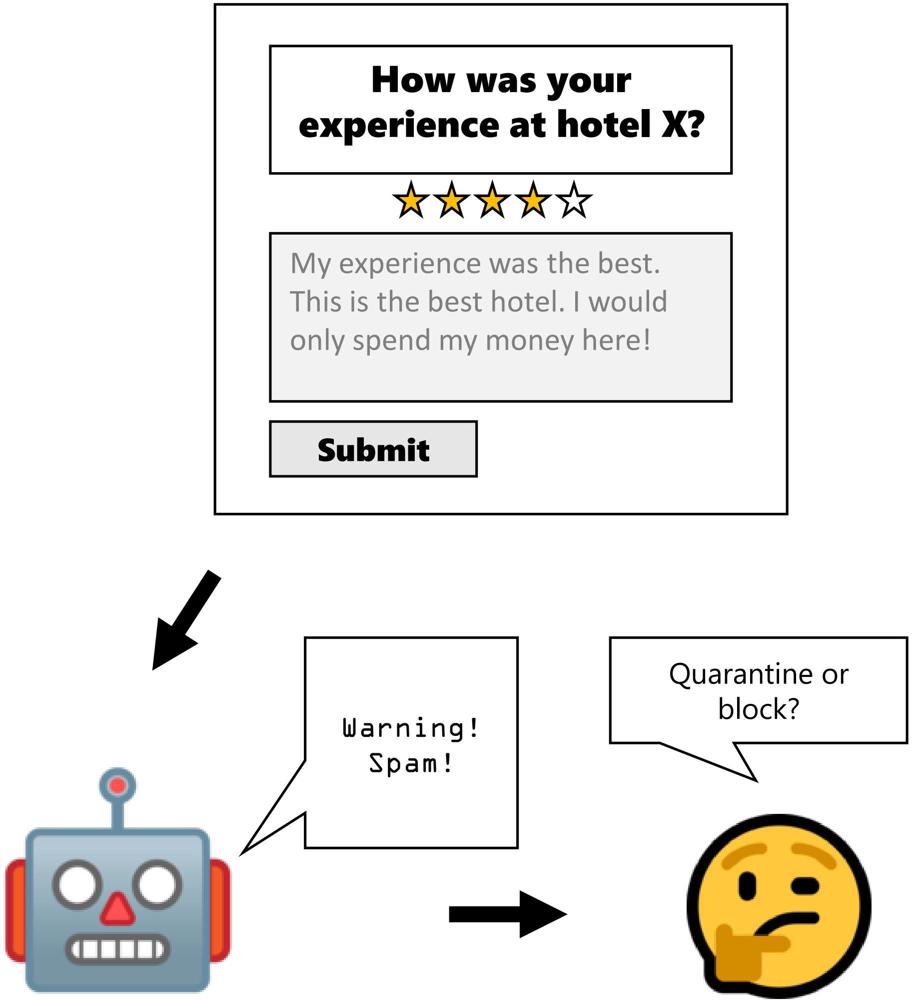
A Hotel Booking site needs to know whether it can detect deceptive reviews within a reasonable level of accuracy and wants to know what to look for in a fake review. This would be the basis of a "first line of defense" against deceptive content.

Goal:

- Build a model to detect deceptive reviews
- Identify the most important characteristics when detecting deceptive reviews

Solution:

Utilize techniques from natural language processing and machine learning to classify whether a hotel review is deceptive or truthful.



Dataset

- 1,600 records
  - 50%: Deceptive
- Deceptive content generated by paid reviewers through Amazon’s Mturk, a market place for work requiring human intelligence.
- Truthful reviews from TripAdvisor
- Data spans 20 Hotels in the Chicago Area

	source				text			
deceptive	deceptive		truthful		deceptive		truthful	
polarity	negative	positive	negative	positive	negative	positive	negative	positive
hotel								
affinia	20	20	20	20	20	20	20	20
allegro	20	20	20	20	20	20	20	20
amalfi	20	20	20	20	20	20	20	20
ambassador	20	20	20	20	20	20	20	20
conrad	20	20	20	20	20	20	20	20
fairmont	20	20	20	20	20	20	20	20
hardrock	20	20	20	20	20	20	20	20
hilton	20	20	20	20	20	20	20	20
homewood	20	20	20	20	20	20	20	20
hyatt	20	20	20	20	20	20	20	20
intercontinental	20	20	20	20	20	20	20	20
james	20	20	20	20	20	20	20	20
knickerbocker	20	20	20	20	20	20	20	20
monaco	20	20	20	20	20	20	20	20
omni	20	20	20	20	20	20	20	20
palmer	20	20	20	20	20	20	20	20
sheraton	20	20	20	20	20	20	20	20
sofitel	20	20	20	20	20	20	20	20
swissotel	20	20	20	20	20	20	20	20
talbott	20	20	20	20	20	20	20	20

Data Prep

- Create 3 Types of Variables:
  1. *Parts of Speech:*
     - Extract components of speech from each sentence. Prep for later steps.
  2. *Polarity and Hotel variables:*
     - Information that came with each review
  3. *Weight Meaningful Words and Group words into "Topics"*
     - Reduce all possible words using Principle Component Analysis

Parts of Speech Variables

Approach and Example:

Original Sentence	"Tokenized" sentence	Remove "Stopwords" (words with little meaning)	Lemmatize (Standardize Tense)	Tag "Parts of Speech"
We stayed for a one night getaway with the family...	[We, stayed, for, a, one, night, getaway, with, the, family, ...]	[We, stayed, one, night, getaway, family, ...]	[We, stay, one, night, getaway, family, ...]	[PRON, VERB, ADP, NUM, NOUN, NOUN, NOUN, ...]

Result

Variables	Description
1	Noun Count
2	Punctuation Count
3	Verb Count
4	Pronoun Count
5	Adjective Count

Categorical Data

Categorical Data → Dummy Variables (22 Variables)

hotel	polarity	source
conrad	positive	TripAdvisor
hyatt	positive	TripAdvisor
hyatt	positive	TripAdvisor
omni	positive	TripAdvisor
hyatt	positive	TripAdvisor

hotel_palmer	hotel_sheraton	hotel_sofitel	hotel_swissotel	hotel_talbott	polarity_negative	polarity_positive
0	0	0	0	0	0	1
0	0	0	0	0	0	1
0	0	0	0	0	0	1
0	0	0	0	0	0	1
0	0	0	0	0	0	1

Transforming our Words to Topics

Step 1. Prepped Data

	I	We	Family	Hotel	Chicago	…	Mth Word
Review 1	0	1	0	1	1	…	1
Review 2	0	0	1	1	1	…	1
Review 3	1	0	1	1	1	…	1
Review 4	1	1	1	0	0	…	0
Review 5	1	0	0	1	0	…	1
…	…	…	…	…	…	…	1
Review N	1	1	1	1	0	1	0

Step 2. Weight according to how often a word appears **within** and **across** reviews

	I	We	Family	Hotel	Chicago	…	Mth Word
Review 1	0	0.061086	0	0.844036	0.12947	…	0.390538
Review 2	0	0	0.563086	0.329746	0.837083	…	0.674166
Review 3	0.048645	0	0.223883	0.001823	0.070719	…	0.782855
Review 4	0.71509	0.25766	0.99956	0	0	…	0
Review 5	0.269133	0	0	0.955859	0	…	0.635496
…	…	…	…	…	…	…	0.125536
Review N	0.981846	0.49423	0.52418	0.76736	0	0.642885	0

Step 3. Group words into "Topics"

	Topic 1	Topic 2	Topic 3	Topic 4	Topic 5	…	Topic N
Review 1	0.128233	0.580455	0.834772	0.350905	0.683768	…	0.939091
Review 2	0.216463	0.477634	0.435082	0.484868	0.069642	…	0.322344
Review 3	0.862888	0.363755	0.583587	0.568093	0.071966	…	0.32307
Review 4	0.488867	0.001899	0.073112	0.943699	0.365091	…	0.911353
Review 5	0.321125	0.893072	0.748353	0.665679	0.235915	…	0.753524
…	…	…	…	…	…	…	0.000744
Review N	0.793765	0.46612	0.820786	0.795149	0.225554	0.932633	0.059631

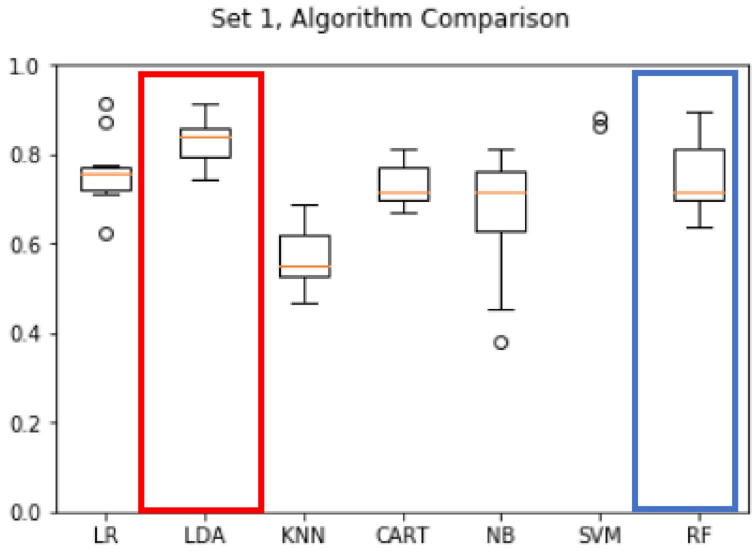
Creating a Baseline

- 7 Classification Models
  - Logistic Regression
  - Linear Discriminant Analysis (LDA)
  - K Nearest Neighbor Classification
  - Decision Trees
  - Naive Bayes (NB)
  - Support Vector Classifier (SVM)
  - Random Forest (RF)

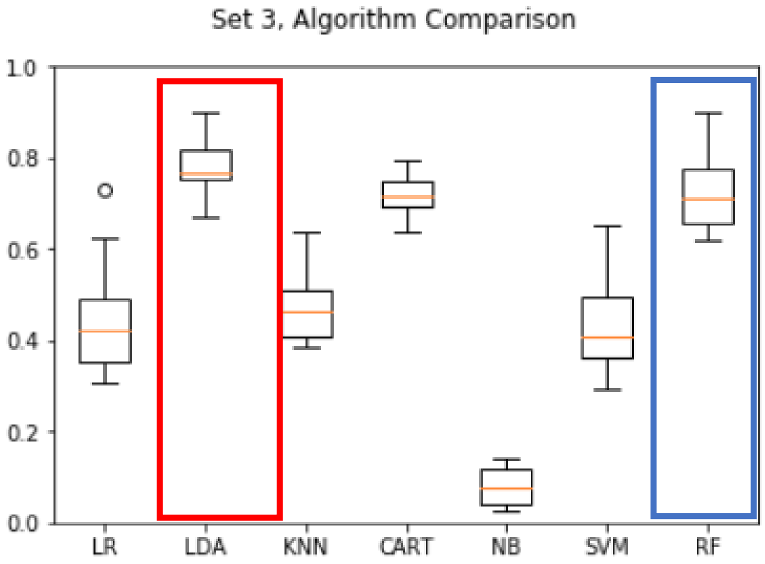
- Three sets of Variables :
  - X1 = Topics Only
  - X2 = Topics + Parts of Speech Metrics
  - X3 = Topics + Parts of Speech Metrics + Dummy Variables

Baseline Accuracy Scores

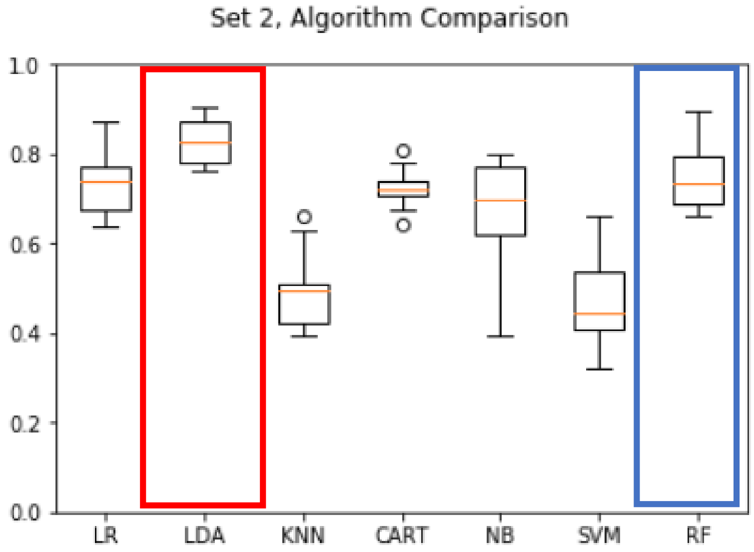
Topics Only



Previous + Parts of Speech Metrics



Previous + Dummy Variables



High Performing Models:
Linear discriminant analysis (Red, 82.8,% 78%, 83%)
Random Forest (Blue 74%, 72%, 75%)

Going beyond baseline

Use Random Forest classifier because…

1. Many variables to data (> 300)
2. Many parameters to be tuned
    1. LDA cannot be "Tuned" for performance
3. Does well with highly correlated variables
4. Easily get variable importance

"Forest" of models

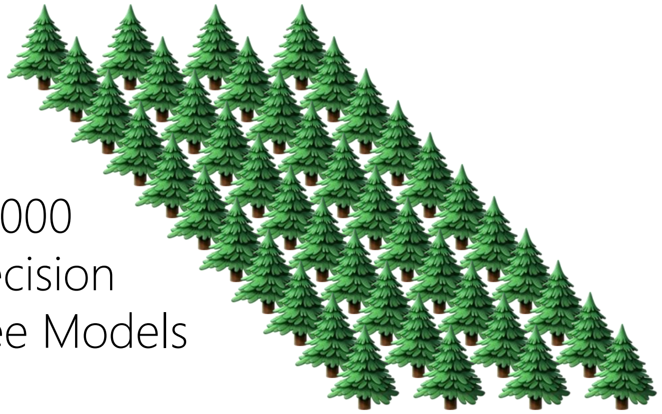
 **=**

100's of Decision Tree models at random subsets of data

Results (1)

- Best Accuracy: 84.6%
  - 1.6 % improvement compared to LDA
  - + 10% improvement compared to untuned Random Forest model

- Details on best parameters:
  - Max Tree Depth = 50
  - Minimum Samples per Node = 50
  - Gini Impurity
  - Of 322 variables, randomly select 18 per tree
  - 1000 Decision Trees
  - 10 folds cross validation



10,000
Decision
Tree Models

Results (2)

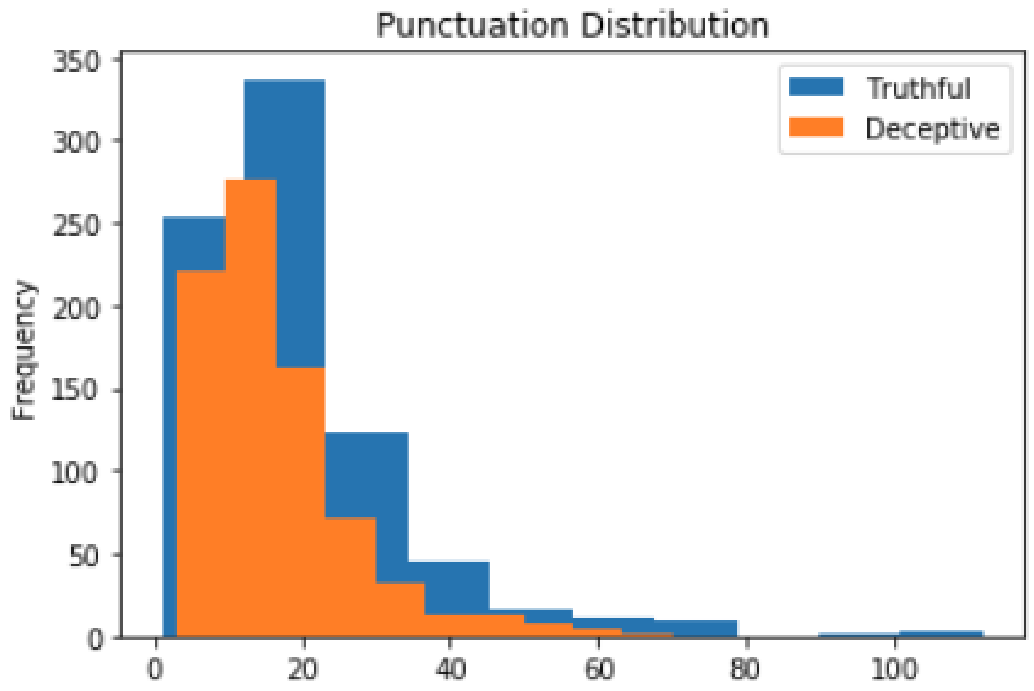
- Topic 2 turned out to be most important. Here are some important words – most of them are positive adjectives and adverbs.

Variable	Importance Rank
Topic 2	1
Topic 6	2
Topic 3	3
Topic 4	4
Topic 12	5
Topic 15	6
Punctuation Count	7

	Word
0	'
1	great
2	comfortable
3	beautiful
4	enjoyed
5	definitely
6	When
7	wonderful
8	called
9	helpful

Results (3)

Descriptively, there is a noticeable difference in punctuation count between truthful and deceptive reviews.



Takeaways

- We ended up with a model that achieved 84.6% accuracy with cross validation
- We have identify a list of important “Topics” and variables.
  - Each topic can be “Opened up” to see what type of words show up.
    - The most important topic includes many key adjectives and adverbs
  - Punctuation frequency is also a good indicator

Takeaways and Next Steps

- Productionizing:
  - Given this reasonable accuracy score, this model can be put into production to act as a "first line of defense" against fake reviews
- Warnings:
  - Generalizability – it is important to continue to test this model for changes in effectiveness that warrant re-training.
- Enrichment:
  - The model can be enriched by further experimenting with different preprocessing step